

11/19/99  
jc714 U.S. PTO

PATENT

Docket No. F9618 A

Box Patent Application  
Assistant Commissioner for Patents  
Washington, D.C. 20231

NEW APPLICATION TRANSMITTAL

Transmitted herewith for filing is the patent application of

Inventor(s): Igor Zlokarnik et al.

jc688 U.S. PTO  
09/443957  
11/19/99

**WARNING:** Patent must be applied for in the name(s) of all of the actual inventor(s). 37 CFR 1.41(a) and 1.53(b).

For (title): Voice-Activated Control For Electrical Device

1. Type of Application

This new application is for a(n) (check one applicable item below):

- ☒ Original  
☐ Design  
☐ Plant

**WARNING:** Do not use this transmittal for a completion in the U.S. of an International Application under 35 U.S.C. 371(c)(4) unless the International Application is being filed as a divisional, continuation or continuation-in-part application.

**NOTE:** If one of the following 3 items apply then complete and attach ADDED PAGES FOR NEW APPLICATION TRANSMITTAL WHERE BENEFIT OF A PRIOR U.S. APPLICATION CLAIMED and a NOTIFICATION IN PARENT APPLICATION OF THE FILING OF THIS CONTINUATION APPLICATION.

- ☐ Divisional  
☐ Continuation  
☐ Continuation-in-part (CIP)

CERTIFICATION UNDER 37 CFR 1.10

I hereby certify that this New Application Transmittal and the documents referred to as enclosed therein are being deposited with the United States Postal Service on this date November 19, 1999, in an envelope as "Express Mail Post Office to Addressee" Mailing Label Number EJ861387775US, addressed to the: Assistant Commissioner for Patents, Washington, D.C. 20231.

Lawrence G. Fridman

(type or print name of person mailing paper)

Signature of person mailing paper

**NOTE:** Each paper or fee referred to as enclosed herein has the number of the "Express Mail" mailing label placed thereon prior to mailing. 37 CFR 1.10(b).

**WARNING:** Certificate of mailing (first class) or facsimile transmission procedures of 37 CFR 1.8 cannot be used to obtain a date of mailing or transmission for this correspondence.

(Application Transmittal [4-1]—page 1 of 5)

## 2. Benefit of Prior U.S. Application(s) (35 USC 120)

NOTE: If the new application being transmitted is a divisional, continuation or a continuation-in-part of a parent case, or where the parent case is an International Application which designated the U.S., then check the following item and complete and attach ADDED PAGES FOR NEW APPLICATION TRANSMITTAL WHERE BENEFIT OF PRIOR U.S. APPLICATION(S) CLAIMED.

- ☐ The new application being transmitted claims the benefit of prior U.S. application(s) and enclosed are ADDED PAGES FOR NEW APPLICATION TRANSMITTAL WHERE BENEFIT OF PRIOR U.S. APPLICATION(S) CLAIMED.

## 3. Papers Enclosed Which Are Required For Filing Date Under 37 CFR 1.53(b) (Regular) or 37 CFR 1.153 (Design) Application

27 Pages of specification

8 Pages of claims

1 Pages of Abstract

6 Sheets of drawing (FIGS 1-8)

- ☒ formal  
☐ informal

WARNING: DO NOT submit original drawings. A high quality copy of the drawings should be supplied when filing a patent application. The drawings that are submitted to the Office must be on strong, white, smooth, and non-shiny paper and meet the standards according to § 1.84. If corrections to the drawings are necessary, they should be made to the original drawing and a high-quality copy of the corrected original drawing then submitted to the Office. Only one copy is required or desired. Comments on proposed new 37 CFR 1.84. Notice of March 9, 1988 (1990 O.G. 57-62).

NOTE: "Identifying indicia such as the serial number, group and unit, title of the invention, attorney's docket number, inventor's name, number of sheets, etc., not to exceed 2 3/4 inches (7.0 cm.) in width may be placed in a centered location between the side edges within three fourths inch (19.1 mm.) of the top edge. Either this marking technique on the front of the drawing or the placement, although not preferred, of this information and the title of the invention on the back of the drawings is acceptable." Proposed 37 CFR 1.84(1). Notice of March 9, 1988 (1090 O.G. 57-62).

## 4. Additional papers enclosed

- ☐ Preliminary Amendment  
☐ Information Disclosure Statement (37 CFR 1.98)  
☐ Form PTO-1449  
☐ Citations  
☐ Declaration of Biological Deposit  
☐ Submission of "Sequence Listing," computer readable copy and/or amendment pertaining thereto for biotechnology invention containing nucleotide and/or amino acid sequence.  
☐ Authorization of Attorney(s) to Accept and Follow Instructions from Representative  
☐ Special Comments  
☐ Other

5. Declaration or oath

☒ Enclosed

executed by (check all applicable boxes)

☐ inventor(s).

☐ legal representative of inventor(s). 37 CFR 1.42 or 1.43

☐ joint inventor or person showing a proprietary interest on behalf of inventor who refused to sign or cannot be reached.

☐ this is the petition required by 37 CFR 1.47 and the statement required by 37 CFR 1.47 is also attached. See item 13 below for fee.

☐ Not Enclosed.

**WARNING:** Where the filing is a completion in the U.S. of an International Application but where a declaration is not available or where the completion of the U.S. application contains subject matter in addition to the International Application the application may be treated as a continuation or continuation-in-part, as the case may be, utilizing ADDED PAGE FOR NEW APPLICATION TRANSMITTAL WHERE BENEFIT OF PRIOR U.S. APPLICATION CLAIMED.

☐ Application is made by a person authorized under 37 CFR 1.41(c) on behalf of all the above named inventor(s). The declaration or oath, along with the surcharge required by 37 CFR 1.16(e) can be filed subsequently.

**NOTE:** It is important that all the correct inventor(s) are named for filing under 37 CFR 1.41(c) and 1.53(b).

☐ Showing that the filing is authorized. (Not required unless called into question. 37 CFR 1.41(d).

6. Inventorship Statement

**WARNING:** If the named inventors are each not the inventors of all the claims an explanation, including the ownership of the various claims at the time the last claimed invention was made, should be submitted.

The inventorship for all the claims in this application are:

☒ The same

or

☐ Are not the same. An explanation, including the ownership of the various claims at the time the last claimed invention was made,

☐ is submitted.

☐ will be submitted.

7. Language

**NOTE:** An application including a signed oath or declaration may be filed in a language other than English. A verified English translation of the non-English language application and the processing fee of \$130.00 required by 37 CFR 1.17(k) is required to be filed with the application or within such time as may be set by the Office. 37 CFR 1.52(d).

**NOTE:** A non-English oath or declaration in the form provided or approved by the PTO need not be translated. 37 CFR 1.69(b).

☒ English

☐ non-English

☐ the attached translation is a verified translation. 37 CFR 1.52(d).

## 8. Assignment

☒ An assignment of the invention to Voice Signal Technologies, Inc.

☐ is (are) attached. A separate "ASSIGNMENT COVER LETTER ACCOMPANYING NEW PATENT APPLICATION" is also attached.

☒ will follow.

NOTE: "If an assignment is submitted with a new application, send two separate letters—one for the application and one for the assignment." Notice of May 4, 1990 (1114 D.G. 77-78).

## 9. Certified Copy

Certified copy(ies) of application(s)

(country)	(appln. no.)	(filed)
(country)	(appln. no.)	(filed)
(country)	(appln. no.)	(filed)

from which priority is claimed

☐ is(are) attached.

☐ will follow.

NOTE: The foreign application forming the basis for the claim for priority must be referred to in the oath or declaration. 37 CFR 1.55(a) and 1.63.

NOTE: This item is for any foreign priority for which the application being filed directly relates. If any parent U.S. application or International Application from which this application claims benefit under 35 U.S.C. 120 is itself entitled to priority from a prior foreign application then complete item 18 on the ADDED PAGES FOR NEW APPLICATION TRANSMITTAL WHERE BENEFIT OF PRIOR U.S. APPLICATION(S) CLAIMED.

## 10. Fee Calculation (37 CFR 1.16)

A. ☐ Regular application

CLAIMS AS FILED			
Number filed	Number Extra	Rate	Basic Fee 37 CFR 1.16(a) \$ 760.00
Total Claims (37 CFR 1.16(c))	26 - 20 = 6	X	\$18.00
Independent Claims (37 CFR 1.16(b))	4 - 3 = 1	X	\$78.00
Multiple dependent claim(s), if any (37 CFR 1.16(d))			\$220.00

☐ Amendment cancelling extra claims enclosed.

☐ Amendment deleting multiple-dependencies enclosed.

☐ Fee for extra claims is not being paid at this time.

NOTE: If the fees for extra claims are not paid on filing they must be paid or the claims cancelled by amendment, prior to the expiration of the time period set for response by the Patent and Trademark Office in any notice of fee deficiency. 37 CFR 1.16(d).

Filing Fee Calculation

\$ 946.00

(Application Transmittal [4-1]—page 4 of 7)

- B. ☒ **Design application**  
(\$330.00—37 CFR 1.16(f))

Filing Fee Calculation \$ \_\_\_\_\_

- C. ☐ **Plant application**  
(\$460.00—37 CFR 1.16(g))

Filing fee calculation \$ \_\_\_\_\_

**11. Small Entity Statement(s)**

- ☒ Verified Statement(s) that this is a filing by a small entity under 37 CFR 1.9 and 1.27 is(are) attached.

Filing Fee Calculation (50% of A, B or C above) \$ 473.00

NOTE: Any excess of the full fee paid will be refunded if a verified statement and a refund request are filed within 2 months of the date of timely payment of a full fee. 37 CFR 1.28(a).

**12. Request for International-Type Search (37 CFR 1.104(d)) (complete, if applicable)**

- ☐ Please prepare an international-type search report for this application at the time when national examination on the merits takes place.

**13. Fee Payment Being Made At This Time**

- ☐ Not Enclosed

- ☐ No filing fee is to be paid at this time. (This and the surcharge required by 37 CFR 1.16(e) can be paid subsequently.)

- ☒ Enclosed

- ☒ basic filing fee \$ 473.00

- ☐ recording assignment  
(\$40.00; 37 CFR 1.21(h)) \$ \_\_\_\_\_

- ☐ petition fee for filing by other  
than all the inventors or person  
on behalf of the inventor where  
inventor refused to sign or cannot  
be reached. (\$130.00; 37 CFR  
1.47 and 1.17(h)) \$ \_\_\_\_\_

- ☐ for processing an application with  
a specification in a non-English  
language. (\$130.00; 37 CFR 1.52(d) and  
1.17(k)) \$ \_\_\_\_\_

- ☐ processing and retention fee  
(\$130.00; 37 CFR 1.53(d) and 1.21(l))

- ☐ fee for international-type search report (\$35.00;  
37 CFR 1.21(e)). \$ \_\_\_\_\_

NOTE: 37 CFR 1.21(l) establishes a fee for processing and retaining any application which is abandoned for failing to complete the application pursuant to 37 CFR 1.53(d) and this, as well as the changes to 37 CFR 1.53 and 1.78, indicate that in order to obtain the benefit of a prior U.S. application, either the basic filing fee must be paid or the processing and retention fee of § 1.21(l) must be paid within 1 year from notification under § 53(d).

**Total fees enclosed**

\$ 473.00

14. Method of Payment of Fees

- ☒ Check in the amount of \$ 473.00
- ☐ Charge Account No. \_\_\_\_\_ in the amount of \$ \_\_\_\_\_. A duplicate of this transmittal is attached.

NOTE: Fees should be itemized in such a manner that it is clear for which purpose the fees are paid. 37 CFR 1.22(b).

15. Authorization to Charge Additional Fees

WARNING: If no fees are to be paid on filing the following items should not be completed.

WARNING: Accurately count claims, especially multiple dependent claims, to avoid unexpected high charges, if extra claim charges are authorized.

- ☐ The Commissioner is hereby authorized to charge the following additional fees by this paper and during the entire pendency of this application to Account No. \_\_\_\_\_:

- ☐ 37 CFR 1.16(a), (t) or (g) (filing fees)
- ☐ 37 CFR 1.16(b), (c) and (d) (presentation of extra claims)

NOTE: Because additional fees for excess or multiple dependent claims not paid on filing or on later presentation must only be paid or these claims cancelled by amendment prior to the expiration of the time period set for response by the PTO in any notice of fee deficiency (37 CFR 1.16(d)), it might be best not to authorize the PTO to charge additional claim fees, except possibly when dealing with amendments after final action.

- ☐ 37 CFR 1.16(e) (surcharge for filing the basic filing fee and/or declaration on a date later than the filing date of the application)
- ☐ 37 CFR 1.17 (application processing fees)

WARNING: While 37 CFR 1.17(a), (b), (c) and (d) deal with extensions of time under § 1.136(a) this authorization should be made only with the knowledge that: "Submission of the appropriate extension fee under 37 C.F.R. 1.136(a) is to no avail unless a request or petition for extension is filed." (Emphasis added). Notice of November 5, 1985 (1060 O.G. 27).

- ☐ 37 CFR 1.18 (issue fee at or before mailing of Notice of Allowance, pursuant to 37 CFR 1.311(b))

NOTE: Where an authorization to charge the issue fee to a deposit account has been filed before the mailing of a Notice of Allowance, the issue fee will be automatically charged to the deposit account at the time of mailing the notice of allowance. 37 CFR 1.311(b).

NOTE: 37 CFR 1.28(b) requires "Notification of any change in loss of entitlement to small entity status must be filed in the application . . . prior to paying, or at the time of paying, . . . issue fee". From the wording of 37 CFR 1.28(b): (a) notification of change of status must be made even if the fee is paid as "other than a small entity" and (b) no notification is required if the change is to another small entity.

16. Instructions As To Overpayment

- ☐ credit Account No. \_\_\_\_\_
- ☐ refund

Reg. No. 31,615

Tel. No. (973) 779-2580

SIGNATURE OF ATTORNEY

Lawrence G. Fridman

Type or print name of attorney

Silber & Fridman

P.O. Address

66 Mount Prospect Avenue

Clifton, New Jersey 07013-1918

☐ **Incorporation by reference of added pages**

*Check the following item if the application in this transmittal claims the benefit of prior U.S. application(s) (including an international application entering the U.S. stage as a continuation, divisional or C-I-P application) and complete and attach the ADDED PAGES FOR NEW APPLICATION TRANSMITTAL WHERE BENEFIT OF PRIOR U.S. APPLICATION(S) CLAIMED*

- ☐ Plus Added Pages For New Application Transmittal Where Benefit Of Prior U.S. Application(s) Claimed

Number of pages added \_\_\_\_\_

- ☐ Plus Added Pages For Papers Referred To In Item 4 Above

Number of pages added \_\_\_\_\_

- ☐ Plus "Assignment Cover Letter Accompanying New Application"

Number of pages added \_\_\_\_\_

☐ **Statement Where No Further Pages Added**

*(If no further pages form a part of this Transmittal then end this Transmittal with this page and check the following item)*

- ☒ This transmittal ends with this page.

**PATENT**

Attorney's Docket No.

**F9618 A**

Applicant or Patentee:

**Igor Zlokarnik et al.**

Serial or Patent No.: 0 /

Filed or Issued:

For: **Voice-Activated Control For Electrical Device**

**VERIFIED STATEMENT (DECLARATION) CLAIMING SMALL ENTITY  
STATUS (37 CFR 1.9(f) and 1.27(c))—SMALL BUSINESS CONCERN**

I hereby declare that I am

- ☐ the owner of the small business concern identified below:
- ☒ an official of the small business concern empowered to act on behalf of the concern identified below:

NAME OF CONCERN

ADDRESS OF CONCERN

I hereby declare that the above identified small business concern qualifies as a small business concern as defined in 13 CFR 121.3-18, and reproduced in 37 CFR 1.9(d), for purposes of paying reduced fees under Section 41(a) and (b) of Title 35, United States Code, in that the number of employees of the concern, including those of its affiliates, does not exceed 500 persons. For purposes of this statement, (1) the number of employees of the business concern is the average over the previous fiscal year of the concern of the persons employed on a full-time, part-time or temporary basis during each of the pay periods of the fiscal year, and (2) concerns are affiliates of each other when either, directly or indirectly, one concern controls or has the power to control the other, or a third-party or parties controls or has the power to control both.

I hereby declare that rights under contract or law have been conveyed, to and remain with the small business concern identified above with regard to the invention, entitled

**Voice-Activated Control For Electrical Device**

by inventor(s) **Igor Zlokarnik et al.**

described in

- ☐ the specification filed herewith.
- ☐ application serial no. 0 / , filed .
- ☐ patent no. , issued .

If the rights held by the above identified small business concern are not exclusive, each individual, concern or organization having rights in the invention is listed below\* and no rights to the invention are held by any person, other than the inventor, who would not qualify as an independent inventor under 37 CFR 1.9(c) if that person made the invention, or by any concern which would not qualify as a small business concern under 37 CFR 1.9(d) or a nonprofit organization under 37 CFR 1.9(e).

*\*NOTE: Separate verified statements are required from each named person, concern or organization having rights to the invention averring to their status as small entities. (37 CFR 1.27).*

NAME \_\_\_\_\_

ADDRESS \_\_\_\_\_

☐ INDIVIDUAL

☐ SMALL BUSINESS CONCERN

☐ NONPROFIT ORGANIZATION

NAME \_\_\_\_\_

ADDRESS \_\_\_\_\_

☐ INDIVIDUAL

☐ SMALL BUSINESS CONCERN

☐ NONPROFIT ORGANIZATION

I acknowledge the duty to file, in this application or patent, notification of any change in status resulting in loss of entitlement to small entity status prior to paying, or at the time of paying, the earliest of the issue fee or any maintenance fee due after the date on which status as a small business entity is no longer appropriate. (37 CFR 1.28(b)).

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application, any patent issuing thereon, or any patent to which this verified statement is directed.

NAME OF PERSON SIGNING Daniel L. Roth

TITLE OF PERSON OTHER THAN OWNER President

ADDRESS OF PERSON SIGNING 432 Columbia Street

Cambridge, MA 02141

SIGNATURE



Date

Nov 15 1999

**VOICE-ACTIVATED CONTROL FOR ELECTRICAL DEVICES**

**FIELD OF THE INVENTION**

This invention relates to the field of speech recognition and, more particularly, to  
5 utilizing human speech for controlling voltage supplied to electrical devices, such as lights,  
lighting fixtures, electrical outlets, volume, or any other electrical device.

**BACKGROUND OF THE INVENTION**

The ability to detect human speech and recognize phonemes has been the subject of a  
great deal of research and analysis. Human speech contains both voiced and unvoiced sounds.  
10 Voiced speech contains a set of predominant frequency components known as formant  
frequencies which are often used to identify a distinct sound.

Recent advances in speech recognition technology have enabled speech recognition  
systems to migrate from the laboratory to many services and products. Emerging markets for  
speech recognition systems are appliances that can be remotely controlled by voice  
15 commands. With the highest degree of consumer convenience in mind, these appliances  
should ideally always be actively listening for the voice commands (also called keywords) as  
opposed to having only a brief recognition window. It is known that analog audio input from  
a microphone can be digitized and processed by a micro-controller, micro-processor, micro-  
computer or other similar devices capable of computation. A speech recognition algorithm  
20 can be applied continuously to the digitized speech in an attempt to identify or match a

speech command. Once the desired command has been found, circuitry which controls the amount of current delivered to a lighting fixture or other electrical device can be regulated in the manner appropriate for the command which has been detected.

One problem in speech recognition is to verify the occurrence of keywords in an unknown speech utterance. The main difficulty arises from the fact that the recognizer must spot a keyword embedded in other speech or sounds ("wordspotting") while at the same time reject speech that does not include any of the valid keywords. Filler models are employed to act as a sink for out-of vocabulary speech events and background sounds.

The performance measure for wordspotters is the Figure of Merit (FOM), which is the average keyword detection rate over the range of 1-10 false alarms per keyword per hour. The FOM increases with the number of syllables contained in a keyword (e.g. Wilcox, L.D. and bush, M.A. : "Training and search algorithms for an interactive wordspotting system" Proc. of ICASSP, Vol. II, pp 97-100, 1992) because more information is available for decision making. While using longer voice commands provides an easy way of boosting the performance of wordspotters, it is more convenient for users to memorize and say short commands. A speech recognition system's susceptibility to a mistaken recognition, i.e. a false alarm, generally decreases with the length of the command word. A longer voice command makes it more difficult for a user to remember the voice command vocabulary, which may have many individual words that must be spoken in a particular sequence.

Some speech recognition systems require the speaker to pause between words, which is known as "discrete dictation." The intentional use of speech pauses in wordspotting is reminiscent of the early days of automatic speech recognition (e.g. Rabiner, L.R.: "On

creating reference templates for speaker-independent recognition of isolated words", IEEE Trans, vol. ASSP-26, no 1, pp. 34-42, February, 1978), where algorithmic limitations required the user to briefly pause between words. These early recognizers performed so-called isolated word-recognition that required the words to be spoken separated by pauses in order to facilitate the detection of word endpoints, i.e. the start and end of each word. One technique for detecting word endpoints is to compare the speech energy with some threshold value and identify the start of the word as the point at which the energy first exceeds the threshold value and the end as the point at which energy drops below the threshold value (e.g. Lamel, L.F. et al: "An Improved Endpoint Detector for Isolated Word Recognition", IEEE Trans., Vol. ASSP-29, pp. 777-785, August, 1981). Once the endpoints are determined, only that part of the input that corresponds to speech is used during the pattern classification process. In this prior art technique, the pause is not analyzed and therefore is not used in the pattern classification process.

Speech Recognition systems include those based on Artificial Neural Networks (ANN), Dynamic Time Warping (DTW), and Hidden Markov Models (HMM).

DTW is based on a non-probabilistic similarity measure, wherein a prestored template representing a command word is compared to incoming data. In this system, the start point and end point of the word is known and the Dynamic Time Warping algorithm calculates the optimal path through the prestored template to match the incoming speech.

The DTW is advantageous in that it generally has low computational and memory requirements and can be run on fairly inexpensive processors. One problem with the DTW is that the start point and the end point must be known in order to make a match to determine

where the word starts and stops. The typical way of determining the start and stop points is to look for an energy threshold. The word must therefore be preceded and followed by a distinguishable, physical speech pause. In this manner, there is initially no energy before the word, then the word is spoken, and then there is no energy after the word. By way of example, if a person were to say <pause> "one" <pause>, the DTW algorithm would recognize the word "one" if it were among the prestored templates. However, if the phrase "recognize the word one now" were spoken, the DTW would not recognize the word "one" because it is encapsulated by other speech. No defined start and end points are detected prior to the word "one" and therefore the speech recognition system can not make any determination about the features of that word because it is encapsulated in the entire phrase. Since it is possible that each word in the phrase has no defined start point and end point for detecting energy, the use of Dynamic Time Warping for continuous speech recognition task has substantial limitations.

In the Artificial Neural Network approach, a series of nodes are created with each node transforming the received data. It is an empirical (probabilistic) technology where some end number of features is entered into the system from the start point and the output becomes the probabilities that those features came from a certain word. One of the major drawbacks of ANN is that it is temporally variable. For example, if a word is said slower or faster than the prestored template, the system does not have the ability to normalize that data and compare it to the data of the stored template. In typical human speech, words are often modulated or vary temporarily, causing problems for speech recognition based on ANN.

The Artificial Neural Network is advantageous in that its architecture allows for a higher compression of templates and therefore requires less memory. Accordingly, it has the ability to compress and use less resources in terms of the necessary hardware than the Hidden Markov Model.

5           The Hidden Markov Model has several advantages over DTW and ANN for speech recognition systems. The HMM can normalize an incoming speech pattern with respect to time. If the templates have been generated at one cadence or tempo and the data comes in at another cadence or tempo, the HMM is able to respond very quickly. For example, the HMM can very quickly adjust for a speaker using two different tempos of the word "run" and "ruuuuuun." 10           Moreover, the HMM processes data in frames of usually (16 to 30 milliseconds), allowing it to have very fast response time. Since each frame is processed in real time, the latency for HMM is less than for DTW algorithms which require an entire segment of speech before processing can begin.

15           Another advantage which distinguishes the HMM over DTW and ANN is that it does not require a defined starting or end point in order to recognize a word. The HMM uses qualitative means of comparing the features in an input stream to the stored templates eliminating the need to distinguish the start and end points. It uses a statistical method to match the sound that is being detected with any sound that is contained in its templates and then outputs a score which is used to determine a match. Although the HMM is superior to 20           its counterparts, it is known from the prior art that its implementation to commercial fixed-point embedded systems ( which are clearly different from PC platforms) has been neglected.

Many prior art speech recognition systems have a detrimental feature with respect to command word template generation. When templates are generated from data produced by recorded human speech, they may not accurately represent the way every person says a command word. For example, if a user's particular speech pattern differs significantly from the template data, then very poor performance from the speech recognition system will be experienced when compared to a user whose speech pattern is more similar to the template data. In an HMM recognizer, words are scored by their probability of occurrence. The closer a word is to its prestored template, the higher its probability is calculated. In order for a word to be considered a match, a preset decision threshold is used. In order to be recognized, the similarity between the uttered word data and the template has to exceed the preset decision threshold. Many speech recognition systems have not provided the user with any means of adjusting the preset decision threshold.

#### SUMMARY OF THE INVENTION

As to one aspect, the invention solves the above-identified problems of the prior art by requiring the user to pause at least in between individual words of an audio or voice command. As an example, the command to turn on the lights becomes "lights<pause>on" or "<pause>lights<pause>on<pause>." Viewing the pauses as a substitute for syllables, this new command exhibits the same number of syllables as "turn lights on" and "please turn the lights on," respectively. This improves the FOM without requiring the speaker to memorize more words and the required word order in a voice command.

Accordingly, it is important to note the following key differences between the use of speech pauses in the present invention and in the prior art isolated-word recognition:

1) The invention treats the speech pauses as part of the keywords and as such treats them just like any other speech sound. Thus, the particular spectral qualities of the input signal during speech pauses are essential for a keyword to be correctly detected. In contrast, the prior art isolated-word recognition discards speech pauses during a pre-processing step; and

2) The purpose of the speech pauses in the present invention is to make the keywords longer rather than to simplify endpoint detection. In fact, no explicit endpoint detection is performed at all in the present invention.

It is therefore an object of the present invention to provide a system and method for more accurately recognizing speech commands without increasing the number of individual command words.

It is a further object of the invention to provide an apparatus for controlling an electrical device, such as a lighting fixture including an incandescent lamp or any other suitable electrical load by speech commands.

It is an even further object of the invention to provide a means for adjusting the threshold comparison value between prestored voice recognition data and uttered audio data to thereby accommodate users with different voice patterns.

According to the invention, an apparatus for voice-activated control of an electrical device has receiving means for receiving at least one audio command generated by a user. The at least one audio command has a command word portion and a pause portion, with each of the audio command portions being at least one syllable in length. Voice recognition data is

provided with a command word portion and a pause portion. Each of the voice recognition data portions are also at least one syllable in length. Voice recognition means is provided for comparing the command word portion and the pause portion of the at least one received audio command with the command word portion and the pause portion, respectively, of the voice recognition data. The voice recognition means generates at least one control signal based on the comparison. Power control means is provided for controlling power delivered to an electrical device. The power control means is responsive to the at least one control signal generated by the voice recognition means for operating the electrical device in response to the at least one audio command generated by the user.

Further according to the invention, a method of activating an electrical device through voice commands, comprises the steps of: recording voice recognition data having a command word portion and a pause portion, each of the voice-recognition data portions being at least one syllable in length; receiving at least one audio command from a user, the at least one audio command having a command word portion and a pause portion, each of the audio command portions being at least one syllable in length; comparing the command word portion and the pause portion of the at least one received audio command with the command word portion and the pause portion, respectively, of the voice recognition data; generating at least one control signal based on the comparison; and controlling power delivered to an electrical device in response to the at least one control signal for operating the electrical device in response to the at least one received audio command.

According to a further embodiment of the invention, an apparatus for voice-activated control of an electrical fixture comprises receiving means for receiving audio data generated

by a user and voice recognition means for determining if the received audio data is a command word for controlling the electrical fixture. The voice recognition means includes a microcontroller with a fixed-point embedded microprocessor, a speech recognition system operably associated with the microcontroller and including a Hidden Markov Model for  
5 comparing data points associated with the received audio data with data points associated with voice recognition data previously stored in the voice recognition means. The voice recognition means generates at least one control signal based on the comparison when the comparison reaches a predetermined threshold value. Power control means are provided for controlling power delivered to the electrical fixture. The power control means is responsive  
10 to the at least one control signal generated by the voice recognition means for operating the electrical device in response to the at least one audio command generated by the user.

According to an even further embodiment of the invention, an apparatus for voice-activated control of an electrical device comprises receiving means for receiving audio data generated by a user and voice recognition means for determining if the received audio data is a  
15 command word for controlling the electrical device. The voice recognition means including a microprocessor for comparing the received audio data with voice recognition data previously stored in the voice recognition means. The voice recognition means generates at least one control signal based on the comparison when the comparison reaches a predetermined threshold value. Power control means is provided for controlling power delivered to the  
20 electrical device. The power control is responsive to the at least one control signal generated by the voice recognition means for operating the electrical device in response to the at least one audio command generated by the user. Also provided is means for adjusting the

predetermined threshold value to thereby cause a control signal to be generated by the voice recognition means when the audio data generated by the user varies from the previously stored voice recognition data.

According to an even further embodiment of the invention, a method of activating an electrical device through an audio command generated by a user comprises recording voice recognition data, receiving audio data generated by a user, comparing the received audio data with the recorded voice recognition data, generating at least one control signal based on the comparison when the comparison reaches a predetermined threshold value, controlling power delivered to an electrical device in response to the at least one control signal to thereby operate the electrical device in response to the generated audio data, and adjusting the predetermined threshold value to generate the control signal when the audio data generated by the user varies from the previously stored voice recognition data.

Other objects, advantages and features of the invention will become apparent upon reading the following detailed description and appended claims, and upon reference to the accompanying drawings.

#### **BRIEF DESCRIPTION OF THE DRAWINGS**

The preferred embodiments of the present invention will hereinafter be described in conjunction with the accompanying drawings, wherein like designations denote like elements throughout the drawings, and wherein:

FIG. 1 shows a block diagram of an apparatus for voice-activated control of an electrical device according to the present invention;

FIG. 2 is an output timing diagram showing the AC output delivered to a lighting fixture or other similar load;

FIG. 3 is a process flow chart for use with the invention for recognizing the presence of a voice command;

5        FIG. 4 is a process flow chart according to the invention for recognizing the presence of a voice command;

FIG. 5 is a chart schematically representing an acceptable energy level of voice command to background noise for actuating an electrical device;

10       FIG. 6 is a chart schematically representing an unacceptable energy level of voice command to background noise for actuating an electrical device;

FIG. 7 is a process flow chart according to a further embodiment of the invention for recognizing the presence of a voice command; and

FIG. 8 is a process flow chart according to an even further embodiment of the invention for recognizing the presence of a voice command.

15

#### **DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS**

Referring now to the drawings, and to FIG. 1 in particular, a functional block diagram of an apparatus 10 for receiving and processing voice commands according to the present invention is illustrated. The apparatus 10 comprises an audio input 12 that is monitored by a micro-controller 14 which performs processing functions associated with a programmed

speech recognition algorithm in an attempt to identify voice commands. Once a voice command has been identified, the micro-controller 14 directs a dimmer control circuit 16 or other electrical control or switching circuit to take an appropriate action. The power required for operation of the micro-controller and other associated electronic units comes from a conventional 5 V DC power supply 18, that may be converted from an alternating current voltage input 20. Once a voice command has been recognized by the micro-controller 14, the control circuit 16 is manipulated to provide appropriate AC (or DC) output 22 to an external electrical device, such as a lamp, electrical outlet, and so on.

The micro-controller 14 preferably includes an 8-bit or 16-bit MCU embedded, fixed-point microprocessor as the central processing unit (CPU), an on-chip analog to digital (A/D) converter, a Read Only Memory (ROM) bank, a Static Random Access Memory (SRAM) bank, and general purpose Input-Output (I/O) ports. This embedded system is specifically different from a PC system, in that it does not have the ability to be modified once the program has been loaded into the chip; it does not have a disk drive or any other means of being changed. It has firmware as opposed to software, and it does not have a monitor or a keypad. It has a very specific purpose that once the system is loaded with the firmware (hardware), it is only allowed to perform one specific task which is to analyze incoming audio data and comparing that data to one or more sets of prestored data to thereby control an outside electrical device. Unlike a PC which has a general purpose processing unit that can do a multitude of tasks from word-processing to game play to speech recognition, for instance; the system of the present invention that runs the Hidden Markov Model is embedded and all contained on a single PC board. The microphone and all of the analog and digital circuitry,

including all of the control circuitry and means for interacting with the speech recognition system is preferably contained on a single printed circuit board. The printed circuit board with all of its components is preferably enclosed in a stand-alone housing (not shown), such as a plastic box, that has a electrical plug for connection with an electrical receptacle. Preferably, the housing also includes electrical receptacles for accepting electrical devices of one type or another to be controlled by voice command.

The audio input 12 may be in the form of a microphone or similar device for inputting audio signals into the apparatus 10. The input signals are then filtered and sent to two amplifiers (not shown) having different gains. The micro-controller processes the analog signal from the channel with the highest gain unless there are data over-ranges. When over-ranges are encountered, the channel with lower gain is utilized. This effectively provides for more dynamic range in the input audio signal without the delay associated with other Automatic Gain Circuits (which have a time delay associated with correction of gain).

An audio output device 24 may be provided as feedback means and instructions to the user. Compressed speech may be stored in System Memory and played back to the user at the appropriate time. This synthesized speech can be used for prompting during a training process where the user(s) is prompted to speak for creating voice templates prior to use. The Audio Output device 24 is preferably in the form of a speaker element.

The apparatus 10 interfaces directly with any source of AC power 20 in general and is specifically adapted to interface with a common household AC (120Vrms in the USA) power outlet via a connector plug (not shown). The connector plug is polarized, with the larger of

the two prongs connecting to "common" (white) of a normal household power outlet. The "hot" (black) input is fused with a 5 amp fuse as a safety measure.

The AC Input 20 is connected to the Dimmer Control Circuit 16 as well as to a full-wave rectifier which is part of the DC Power Supply 18.

5 The Dimmer Control Circuit 16 controls the current delivered to a lighting fixture (not shown) or other similar load and is regulated by a TRIAC. In this manner, it is possible to turn the fixture on or off, so it can be dimmed without dissipative power losses. The TRIAC is driven by an opto-coupler and is therefore electrically isolated from the circuitry of the apparatus 10 which contains the audio input amplifier, the micro-controller 14, and other related functional circuitry.

10 The input AC signal feeds into a zero-crossing detection circuit (not shown) and generates a trigger pulse which is used as the clock input for several one-shots with time constants adjusted to provide pulse trains at 33% and 66% of the AC line frequency duty cycle. These pulse trains are supplied to a multiplexer (not shown) which is controlled by the micro-controller 14 adapted for selecting an appropriate digital pulse train to drive the opto-coupler.

15 Several bits from Port O of the speech recognition chip (bits 2 and 3) select the appropriate channel of the multiplexer which is adapted for driving this circuit. It has four states:

Bit 1	Bit 0	State
0	0	ON
1	0	2/3 ON (dim)
0	1	1/3 ON (low dim)

1	1	OFF
---	---	-----

For the 1/3 ON dim selection, the TRIAC is on 33% of the time, allowing current to flow to the lighting fixture or other similar load. When the TRIAC turns off, the current does not flow to the fixture. This causes the lighting fixture to appear dim. In this technique, power is not dissipated by the dimmer circuit when the lights are dimmed. Output waveforms for each of the above states are illustrated in FIG. 2.

The AC input 20 is connected to the DC power supply 18, which includes a full wave rectifier, voltage regulator, and other components form part of the power supply 18 and provide a 5 Volt supply with minimum ripple (5 millivolts maximum ripple). De-coupling capacitors are preferably used between the power and ground pins of the integrated circuit components to limit noise.

Thus, the invention provides an apparatus which utilizes a microphone for continuously monitoring an audio input signal, and by using speech recognition techniques, identifies the desired human speech commands and controls the current delivered to a lighting fixture including an incandescent lamp or any other suitable load once the appropriate command has been detected.

Turning now to FIG. 3, a typical method of receiving sound data and recognizing voice commands from the sound data for use in the apparatus 10 is illustrated. When the apparatus 10 is in operation, as represented by block 120, and after it has been trained with voice commands for each user, sound data is constantly monitored at block 122 to determine if a voice command has been uttered. The sound data may contain a combination of voice commands and background noise such as voice conversations, machinery, television, radio,

and the like, or any combination thereof. It is not only necessary to separate the voice commands from the background noise in order to issue a control command to a switching device (such as the dimmer control circuit 16 in FIG. 5), but it is also necessary to properly interpret the voice commands received.

5           Speech recognition is a process by which one or more unknown speech utterances are identified. Speech recognition is generally performed by comparing the features of an unknown utterance with the features of known words. Known words as used herein include, without limitation, words, phrases, phonetic units, and/or phonemic units. The features, or characteristics, of the known words are typically defined through a training process wherein  
10       samples of known words are examined and their features are recorded as recognition models in a recognition database. Each recognition model is essentially a reference pattern which represents a single word. Thus, depending on the number of words in a voice command, there will be a corresponding number of recognition models.

15           The sounds in the data stream are classified and ordered at block 124, in a well-known manner. The particular features associated with the unknown utterance are often referred to as a "test pattern." The micro-controller 14 compares the test pattern to one or more voice recognition models or templates 128, 130 and 132, e.g. the trained voice commands, or combinations thereof, as represented by block 126. As illustrated, the templates 128, 130 and 132 might include the voice commands "lightson", "lightsoff", and "dimlights", respectively.  
20       A scoring technique is then used at block 134 to provide a relative measure of how well various recognition model combinations match the test pattern. The unknown utterance is

recognized by the known words associated with the recognition model(s) with which the test pattern most closely matches.

As set forth previously, there are many types of speech recognizers, such as, for example, conventional template-based and Hidden Markov Model (HMM) recognizers, as well as recognizers utilizing recognition models based on neural networks. Without loss of generality, the present invention will be illustrated by way of example to the HMM recognizers.

Wordspotter algorithms typically use a recognition network which allows the test pattern to be recognized in terms of a sequence of keyword, filler and silence models. The keyword model refers to voice command words that are to be recognized by the micro-controller 14, while the filler model refers generally to extraneous background noise that is discarded by the micro-controller. The silence model represents the stationary background noise that occurs before and after keyword utterances. Because of the stationary condition, the silence model is typically modeled with a single looped HMM state in order to detect the beginning and ending of a keyword utterance. In HMM's, a state can be traversed one frame at a time wherein a frame update typically occurs every 10 msec. A formal technique for finding the single best model sequence exists, and is called the Viterbi algorithm (Forney, Jr., G. D., "The Viterbi algorithm", Proc. IEEE, Vol. 61, pp. 268-278, March 1978). Rather than calculating the score for every possible model sequence, the Viterbi algorithm reduces the computational load by eliminating all but the best of those partial sequences that merge at a common point, otherwise known as a recombination point.

The keyword detection and false alarm rate for each keyword can be adjusted by adding a score on entering the keyword model, also known as a word entrance penalty. A higher word-entrance penalty results in a lower key word detection and thus results in a lower false alarm rate. This is because the higher cumulative score makes the keyword model less likely to win against the other models at the recombination point. For example, referring to Fig. 3, when a user utters the command "Dim Lights," the cumulative score  $S_A$  for "Dim" and "Lights" will most likely be lower when compared to the "DIMLIGHTS" template 132 than the cumulative scores for the other templates 128, 130.

After the score  $S_A$  has been generated for the closest matching command template (block 134), the classified data or "test pattern" is compared at block 136 to an "All Other Sounds" template or filler model 138 which has no defined structure, as in background noise such as voice conversations, machinery, television, radio, and the like, or any combination thereof. A score  $S_B$  is then generated at block 140. The scores  $S_A$  and  $S_B$  are then compared at block 142. If the score  $S_A$  is smaller than the score  $S_B$ , then the likelihood that the sound data is a voice command is greater than the likelihood that the sound data is just background noise. Consequently, the system is appropriately triggered at block 144, depending on the closest matching command template in order to control a lamp or other electrical device. If, however the score  $S_A$  is greater than the score  $S_B$ , then the likelihood that the sound data is a voice command is less likely that the sound data is just background noise. Consequently, the apparatus 10 continues to receive sound data at block 122.

With reference now to FIG. 4, a method of receiving sound data and recognizing voice commands from the sound data for use in the apparatus 10 is illustrated. When the apparatus

10 is in operation, as represented by block 150, and after it has been trained with voice commands for a user, sound data is constantly monitored at block 152 to determine if a voice command has been uttered. As described above, the sound data may contain a combination of voice commands and background noise such as voice conversations, machinery, television, radio, and the like, or any combination thereof. The sounds in the data stream are then classified and ordered into a test pattern at block 154, in a well-known manner. The micro-controller 14 compares the test pattern to one or more voice recognition models or templates 158, 160 and 162, e.g. the trained voice commands, or combinations thereof, as represented by block 156.

As illustrated, each of the voice templates 158, 160 and 162 includes a pause model 164 of predefined duration between keywords in the voice commands. Thus, the command "lightson" in the previous example becomes "lights<pause>on". The command "lightsoff" becomes "lights<pause>off", and the command "dimlights" becomes "dim<pause>lights". The duration of the pause model 164 between each command word may vary depending on the particular speaking style of the user(s), but should be at least one syllable (about 200 msec.) in length. As noted above, the Figure of Merit (FOM) increases with the number of syllables contained in a voice command. For an even greater increase in command detection accuracy, a pause may also be added before and/or after each command word. Thus, instead of a two-syllable command for "lightson", a three-syllable command "lights<pause>on" increases the FOM, while a five-syllable command "<pause>lights<pause>on<pause>" greatly increases the FOM without increasing the number of words in the voice command. This is especially advantageous since the user is not required to memorize long phrases (e.g. five-syllable

phrases) for each voice command in order to obtain greater detection accuracy over the prior art.

In order to impose a minimum pause duration of about 200 msec, the pause model 164 needs to contain at least N silence states (represented by  $s_i$  in FIG. 4) where

5 
$$N = \text{minimum pause duration} / \text{frame update} = 200 \text{ msec} / 10 \text{ msec} = 20.$$

Because each state  $s_i$  is modeling the same features as the single silence state, the pause model is created by simply concatenating N silence states. While the minimum duration spent in the pause model is controlled by the number of states, a maximum model duration can be accounted for by adding a pause score whenever a state loops back on itself. The score is called loop transition penalty. Although loop transition penalties cannot completely prevent the best state sequence from remaining longer than a fixed amount of frames in the model, pauses longer than N frames become less likely with duration. Note that each pause model can have a different number of states in order to allow modeling speech pauses of different duration. In particular, pauses at the beginning and end of a voice command do not impede fluency and thus may be chosen to be substantially longer than pauses in between words.

The presence of a pause is preferably determined by analyzing both the spectral content and the energy content of the pause before and/or after the detection of a keyword, depending on the particular sequence of pauses and keywords. If dynamic spectral activity is present at a position in the voice data where the pause should be, such as in the case of voice data, and if the dynamic spectral activity has an energy content that is within a preset energy range of the keyword energy content, then it is determined that no pause has occurred. In the

case where the pause has dynamic spectral activity below the preset energy range, such as in the case of background noise present between keyword utterances, then it is determined that a pause has occurred.

Thus at block 166, a cumulative score  $S_A$  is calculated based on each of the command words and pauses in the voice command to thereby provide a relative measure of how well various recognition model combinations match the test pattern. The unknown utterance must not only have the sound sequence correct, but must also have the unnatural breaks in sound (the pauses) at the correct time in order to create a competitive score for triggering the system.

After the score  $S_A$  has been generated for the closest matching command template, the classified data or "test pattern" is compared at block 168 to the "All Other Sounds" template or filler model 138. A score  $S_B$  is then generated at block 170. The scores  $S_A$  and  $S_B$  are then compared at block 172. If the score  $S_A$  is smaller than the score  $S_B$ , then the likelihood that the sound data is a voice command is greater than the likelihood that the sound data is just background noise. Consequently, the system is appropriately triggered at block 174, depending on the closest matching command template. If, however the score  $S_A$  is greater than the score  $S_B$ , then the likelihood that the sound data is a voice command is less than the likelihood that the sound data is just background noise. Consequently, the apparatus continues to receive sound data at block 152.

Because the invention treats the speech pauses as part of the keywords in the voice commands, a high likelihood score during one or more pauses may compensate the effect of low likelihood scores during the actual words. Thus, some similar sounding utterances may

get accepted as keywords primarily due to a good fit during the speech pauses. In order to prevent this from happening, another recognition network may be used in which the contributions from the pause models and the speech models to the overall likelihood score  $S_A$  can be uncoupled by making the filler and silence models compete with each pause model and speech model individually. This architecture allows speech utterances to be accepted as valid keywords only if each pause and speech model has been determined to be more likely than any sequence of filler and silence models.

In this recognition network, each individual pause and speech model is assigned a word-entrance penalty whose value is made dependent on the best preceding model as determined by the Viterbi recombination. The word-entrance penalty is assigned a finite value, whenever the best predecessor model (as determined by the Viterbi algorithm) corresponds to the syntactically proper predecessor model (as determined by the structure of the keyword model, e.g. (<pause>speech1<pause>speech2< ...)). In all other cases, the word-entrance penalty is assigned an infinite value which will inhibit all remaining parts of the keyword model from being further traversed. A keyword is detected as soon as the last model of that keyword survives the recombination step.

The particular values for the finite word-entrance penalties determine the detection and false alarm rate of the corresponding keywords. Because all possible state sequences through the pause models represent a subset of all state sequences through the filler and silence models, the pause models would always lose during recombination. In order to prevent this from happening, the pause models are rewarded upon entry by using negative (rather than positive) word-entrance penalties.

With reference now to FIG.'s 5 to 7, a further embodiment of the invention is illustrated, wherein like elements in the previous embodiment are represented by like numerals. Additional accuracy of voice command detection can be obtained by comparing the energy ( $E_B$ ) of background noise 180 to the energy ( $E_C$ ) of a keyword utterance 182 (see FIG.'s 5 and 6). If, after it is determined that  $S_A$  is greater than  $S_B$  at block 170 in FIG. 7, the signal strength  $E_C$  is analyzed at block 190 and the background noise signal strength  $E_B$  is analyzed at block 192. It is then determined if the difference  $E$  between the energy  $E_C$  of the keyword utterance and the energy  $E_B$  of the background noise is above a predetermined value at block 194, and as shown in FIG. 5. If  $E$  is above the predetermined value, then the system is triggered at block 196. However, if the sound data does not contain enough energy to meet the established energy difference  $E$  (see FIG. 6), then the micro-controller assumes that the whole sound data is background noise and does not trigger the system. Instead of taking the difference between the energies, the ratio or some other means of comparing the energies can be provided.

With the above-described arrangement, the user can enter commands even in loud environments by talking louder than the background. In this way, the keyword<pause>keyword structure can be maintained even if the pause portion is not actually silent. Consequently, the pause portion is only required to have an energy level that is a certain amount lower than the keywords. In many instances, the pause portion will most likely have an energy level equal to the background noise, unless a noise-canceling microphone or the like is used.

Although it is preferred that the signal strength be analyzed and compared after the voice command has met the criteria for both the command words and the pauses, analysis of the voice commands may be triggered only when the difference or ratio between a detected energy level of sound data and background noise is above a predetermined amount.

5           Turning now to FIG. 8, a modified method of receiving sound data and recognizing voice commands from the sound data according to a further embodiment is illustrated, wherein like parts in the FIG. 3 embodiment are represented by like numerals. In this embodiment, the classified sound data is compared to the word templates at block 126 and scores are generated for all command word templates at block 200. It is then determined  
10   at block 202 which of all the generated scores is the best, i.e. which template matches most closely with the classified sound data. The speech recognition system has an analysis phase that takes the intrinsic features of a speech signal and extracts some data points from that signal. Those data points are then compared to corresponding data points in the prestored templates. The speech recognition system attempts to determine how likely it is that each one  
15   of those data points (in the actual signal) is to the data point that is expected to be seen in the prestored template. Thus, the analysis looks at the incoming data and determines the likelihood of this data correlating with the prestored data. The output is represented as a series of scores or emission probabilities, which represents how close or how well the incoming data and the template data match. As the analyzing portion of the speech  
20   recognition system makes it's comparison, it determines the likelihood of how well a data point "A" in the incoming data will match with data point "A" in each of the templates, the likelihood of how well data point "B" in the incoming data will match with data point "B" in

each of the templates, and so on. Thus, if there are 20 templates in a system, the system will look at the incoming signal and determine its sum probability close to the data points A, B, C, etc., with the corresponding data points in templates 1, 2, 3, ...20. The template with the greatest sum probability is then chosen as the most likely candidate for the uttered command word. For example, the system will indicate that out of 20 prestored templates, the incoming data most highly correlates to Template No. 5. The scores for the templates are compared to the scores of the All Other Sounds Template, as represented by block 204. If the scores of the All Other Sounds Template minus the scores of the most likely template is less than a threshold value  $S$ , then the activation system is triggered, as represented by block 144. If however the difference between the scores is greater than the threshold value  $S$ , then the speech recognition system determines that the incoming data is not one of the 20 templates. It is not a recognized voice command, and the activation system is not triggered.

According to a unique feature of the invention, the threshold value  $S$  can be set by a user through the provision of a threshold adjusting means 206 connected to the micro-controller so that a user can adjust a desired threshold value to thereby effect the outcome of the system. The effect of this is to allow someone whose speech is less compatible with the data represented by templates to use the system with a much higher degree of accuracy or responsiveness by adjusting characteristics that are used to determine whether an uttered word matches the stored data. This is different from speaker adaptive systems used in prior art, wherein the templates were modified based on some user input. In this case, the templates are not adjusted, rather the predetermined threshold used to qualify whether a word is a match is adjustable by the user.

In the invention, preferably, the means for adjusting the threshold value comprises a trim potentiometer that has a digital interface to the speech recognition system. Thus, the setting of the trim potentiometer is sampled or sensed by the speech recognition system. The threshold value  $S$  can be either a discrete step-wise representation of values like integers, such as integers 1, 2, 3, 4, 5, 6, etc. or a continuous relatively fine adjustment, e.g. any value between 0 and 9 rather than the integers. This provides either a fixed step-wise adjustment or continuous adjustment of the threshold value. Although a trim potentiometer is preferred, the threshold adjusting means can take the form of a thumbwheel or sliding potentiometer, a rotary or slide switch with multiple detent switch positions, a DIP switch, an optical encoder, a variably adjustable capacitor, or any other electronic or digital means for adjusting the threshold value.

The threshold adjusting means, no matter what form it takes, sets the point at which the speech recognition system recognizes an uttered command word, i.e. to set the degree of correlation between the incoming data and the template data to thereby trigger the system. This adjusting means allows selective adjustment by a user of the parameters of the similarity measurement, which could either be statistical or rule-based. However, in either case, this would allow a user to adjust a parameter in an embedded system that would normally be fixed. The parameter can have either the effect of adding weight to the output probabilities of the command words or the output probabilities of the All Other Sounds Template. The effect of the adjustment would be to either change the number of false activation's or change the number of positive detections that a speech recognition system makes. Thus, when a stream of incoming data is compared to the data of the prestored templates, adjusting the threshold

value S to a greater value permits a person to utter a command word that is less correlated with the templates in order for the speech recognition system to consider it a match. A sequence is generated that best describes the signal. That sequence of features is then compared to a series of features that stored in the templates. Each data point in the incoming  
5 signal is compared to a corresponding area of each stored template and scores are generated to represent how closely the input signal matches each one of the templates. Accordingly, data points that at one point can be rejected because the All Other Sounds Template have created the best score, will now be considered as a match because the All Other Sounds Template or the other command word templates were adjusted through the threshold value  
10 adjusting means. Although described specifically for use with the Hidden Markov Model, it is to be understood that the threshold adjusting means can be applied to other speech recognition systems.

While the invention has been taught with specific reference to the above-described embodiments, those skilled in the art will recognize that changes can be made in form and  
15 detail without departing from the spirit and the scope of the invention. Thus, the described embodiments are to be considered in all respects only as illustrative and not restrictive. The scope of the invention is, therefore, indicated by the appended claims rather than by the foregoing description. All changes that come within the meaning and range of equivalency of the claims are to be embraced within their scope.

## CLAIMS

The embodiments for which an exclusive property or privilege is claimed are defined as follows:

1           1.     An apparatus for voice activated control of an electrical device, the apparatus  
2     comprising:  
3                 receiving means for receiving at least one audio command generated by a user,  
4     the at least one audio command having a command word portion and a pause portion, each of  
5     the audio command portions being at least one syllable in length;  
6                 voice recognition data having a command word portion and a pause portion,  
7     each of the voice recognition data portions being at least one syllable in length;  
8                 voice recognition means including a Hidden Markov Model for comparing said  
9     command word portion and said pause portion of said at least one received audio command  
10    with said command word portion and said pause portion, respectively, of said voice  
11    recognition data, said voice recognition means generating at least one control signal based on  
12    said comparison; and  
13                 power control means for controlling power delivered to an electrical device,  
14    said power control means being responsive to said at least one control signal generated by said  
15    voice recognition means for operating the electrical device in response to said at least one  
16    audio command generated by the user.

1           2.     The apparatus of claim 1, and further comprising means for analyzing the  
2     pause portion of the received audio command for spectral content, and further wherein said

voice recognition means prevents operation of the electrical device when the spectral content is dynamic.

3. The apparatus of claim 1, wherein said receiving means receives background noise data in conjunction with said audio command, and further comprising means for generating a command word score and a background noise score based on the comparison of the received audio command to the voice recognition data and the background noise data, respectively, said voice recognition means generating said at least one control signal when said command word score exceeds said background noise score.

4. The apparatus of claim 3, and further comprising:  
means for analyzing the command word portion of the received audio command and the background noise data for energy content; and  
means for comparing the energy content of the command word portion to the energy content of the background noise data and generating a corresponding energy comparison value;  
wherein said voice recognition means prevents the generation of said at least one control signal when said energy comparison value is below a predetermined level.

5. The apparatus of claim 1, wherein said receiving means receives background noise data in conjunction with said audio command, and further comprising:  
means for analyzing the command word portion of the received audio command and the background noise data for energy content; and

5 means for comparing the energy content of the command word portion to the  
6 energy content of the background noise data and generating a corresponding energy  
7 comparison value;

8 wherein said voice recognition means prevents the generation of said at least  
9 one control signal when said energy comparison value is below a predetermined level.

1 6. The apparatus of claim 1, wherein each of said at least one audio command and  
2 said voice recognition data comprises at least first and second command word portions  
3 separated by said pause portion and further comprising a second pause portion having at least  
4 one syllable in duration before said first command word portion and a third pause portion  
5 having at least one syllable in duration after said second command word portion.

1 7. The apparatus of claim 1, wherein the voice recognition means further  
2 including a microcontroller with a fixed-point embedded microprocessor, the microprocessor  
3 is chosen from the group of 8-bit and 16-bit MCU microprocessors.

1 8. A method of activating an electrical device through at least one audio  
2 command from a user, the method comprising the steps of:  
3 recording voice recognition data having a command word portion and a pause  
4 portion, each of the voice-recognition data portions being at least one syllable in length;  
5 receiving at least one audio command from a user, the at least one audio  
6 command having a command word portion and a pause portion, each of the audio command  
7 portions being at least one syllable in length;

8 comparing said command word portion and said pause portion of said at least  
9 one received audio command with said command word portion and said pause portion,  
10 respectively, of said voice recognition data;  
11 generating at least one control signal based on said comparison; and  
12 controlling power delivered to an electrical device in response to said at least  
13 one control signal for operating the electrical device in response to said at least one received  
14 audio command.

1 9. The method of claim 8, wherein the step of recording voice recognition data  
2 includes recording the voice of a user while the user utters said at least one audio command.

1 10. The method of claim 8, and further comprising the steps of:  
2 analyzing the pause portion of the received audio command for spectral  
3 content; and  
4 preventing operation of the electrical device when the spectral content is  
5 dynamic.

1 11. The method of claim 8, and further comprising the steps of:  
2 receiving background noise data in conjunction with receiving said at least one  
3 audio command;  
4 comparing the background noise data to the at least one received audio  
5 command;  
6 generating a command word score from the comparison of the received audio  
7 command to the voice recognition data;

8 generating a background noise score based on the comparison of the received  
9 audio command to the background noise data; and  
10 generating said at least one control signal when said command word score  
11 exceeds said background noise score.

1 12. The method of claim 11, and further comprising:  
2 ascertaining a first energy content for the command word portion of the  
3 received audio command;  
4 ascertaining a second energy content for the received background noise data;  
5 comparing the first and second energy contents and generating an energy  
6 comparison value; and  
7 preventing the generation of said at least one control signal when said energy  
8 comparison value is below a predetermined level.

1 13. The method of claim 11, wherein the step of generating a command word  
2 score includes assigning a word entrance penalty to the command word portion and the pause  
3 portion.

1 14. The method of claim 13, wherein the word entrance penalty of the pause  
2 portion is negative.

1 15. The method of claim 8, and further comprising:  
2 receiving background noise data in conjunction with receiving said at least one  
3 audio command;

ascertaining a first energy content for the command word portion of the  
received audio command;  
ascertaining a second energy content for the received background noise data;  
comparing the first and second energy contents and generating a corresponding  
energy comparison value; and  
preventing the generation of said at least one control signal when said energy  
comparison value is below a predetermined level.

16. The apparatus of claim 8, wherein each of said at least one audio command and  
said voice recognition data comprises at least first and second command word portions  
separated by said pause portion.

17. The apparatus of claim 14, and further comprising a second pause portion  
having at least one syllable in duration before said first command word portion and a third  
pause portion having at least one syllable in duration after said second command word  
portion.

18. An apparatus for voice-activated control of an electrical fixture, the apparatus  
comprising:

receiving means for receiving audio data generated by a user;  
voice recognition means for determining if the received audio data is a  
command word for controlling the electrical fixture, the voice recognition means including a  
microcontroller with a fixed-point embedded microprocessor, a speech recognition system  
operably associated with the microcontroller and including a Hidden Markov Model for  
comparing data points associated with the received audio data with data points associated

9 with voice recognition data previously stored in the voice recognition means, said voice  
10 recognition means generating at least one control signal based on said comparison when said  
11 comparison reaches a predetermined threshold value; and

12 power control means for controlling power delivered to the electrical fixture,  
13 said power control means being responsive to said at least one control signal generated by said  
14 voice recognition means for operating the electrical fixture in response to said at least one  
15 audio command generated by the user.

1 19. The apparatus of claim 18, wherein the microprocessor is chosen from the  
2 group of 8-bit and 16-bit MCU microprocessors.

1 20. The apparatus of claim 19, wherein the microcontroller further includes an  
2 analog to digital (A/D) converter for transforming the received voice data to digital data, a  
3 Read Only Memory (ROM) bank, a Static Random Access Memory (SRAM) bank, and  
4 general purpose Input-Output (I/O) ports.

1 21. The apparatus of claim 19, and further comprising a printed circuit board, the  
2 receiving means, the microcontroller, and the power control means being mounted on the  
3 printed circuit board.

1 22. The apparatus of claim 18, wherein the electrical fixture is a lighting fixture,  
2 and the power control means controls power delivered to said lighting fixture.

1 23. The apparatus of claim 22, wherein the power control means includes a control  
2 circuit for turning on, dimming, and turning off said lighting fixture in response to different  
3 audio data uttered by a user.

1           24.    The apparatus of claim 18, and further comprising:

2                   means for adjusting the predetermined threshold value to thereby cause a  
3           control signal to be generated by said voice recognition means when the audio data generated  
4           by the user varies from the previously stored voice recognition data.

1           25.    An apparatus for voice-activated control of an electrical device, the apparatus  
2           comprising:

3                   receiving means for receiving audio data generated by a user;

4                   voice recognition means for determining if the received audio data is a  
5           command word for controlling the electrical device, the voice recognition means including a  
6           microprocessor for comparing the received audio data with voice recognition data previously  
7           stored in the voice recognition means, said voice recognition means generating at least one  
8           control signal based on said comparison when said comparison reaches a predetermined  
9           threshold value;

10                  power control means for controlling power delivered to the electrical device,  
11           said power control means being responsive to said at least one control signal generated by said  
12           voice recognition means for operating the electrical device in response to said at least one  
13           audio command generated by the user; and

14                  means for adjusting the predetermined threshold value to thereby cause a  
15           control signal to be generated by said voice recognition means when the audio data generated  
16           by the user varies from the previously stored voice recognition data.

1           26.    The apparatus of claim 25, wherein the microprocessor is chosen from the  
2           group of 8-bit and 16-bit embedded MCU microprocessors.

## ABSTRACT OF THE DISCLOSURE

AN APPARATUS FOR VOICE-ACTIVATED CONTROL OF AN ELECTRICAL  
DEVICE COMPRISES A RECEIVING ARRANGEMENT FOR RECEIVING AUDIO  
DATA GENERATED BY USER. A VOICE RECOGNITION ARRANGEMENT IS  
5 PROVIDED FOR DETERMINING WHETHER THE RECEIVED AUDIO DATA IS A  
COMMAND WORD FOR CONTROLLING THE ELECTRICAL DEVICE. THE VOICE  
RECOGNITION ARRANGEMENT INCLUDES A MICROPROCESSOR FOR  
COMPARING THE RECEIVED AUDIO DATA WITH VOICE RECOGNITION DATA  
PREVIOUSLY STORED IN THE VOICE RECOGNITION ARRANGEMENT. THE  
10 VOICE RECOGNITION ARRANGEMENT GENERATES AT LEAST ONE CONTROL  
SIGNAL BASED ON THE COMPARISON WHEN THE COMPARISON REACHES A  
PREDETERMINED THRESHOLD VALUE. A POWER CONTROL CONTROLS POWER  
DELIVERED TO THE ELECTRICAL DEVICE. THE POWER CONTROL IS  
RESPONSIVE TO AT LEAST ONE CONTROL SIGNAL GENERATED BY THE  
15 VOICE RECOGNITION ARRANGEMENT FOR OPERATING THE ELECTRICAL  
DEVICE IN RESPONSE TO THE AT LEAST ONE AUDIO COMMAND GENERATED  
BY THE USER. AN ARRANGEMENT FOR ADJUSTING THE PREDETERMINED  
THRESHOLD VALUE IS PROVIDED TO CAUSE A CONTROL SIGNAL TO BE  
GENERATED BY THE VOICE RECOGNITION ARRANGEMENT WHEN THE AUDIO  
20 DATA GENERATED BY THE USER VARIES FROM THE PREVIOUSLY STORED  
VOICE RECOGNITION DATA.

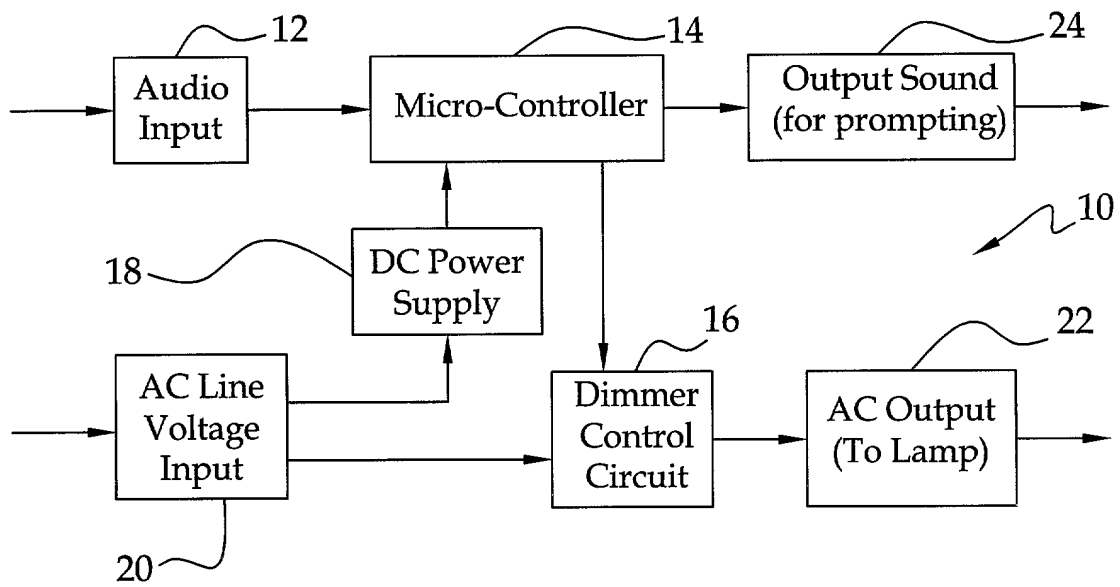


FIG. 1

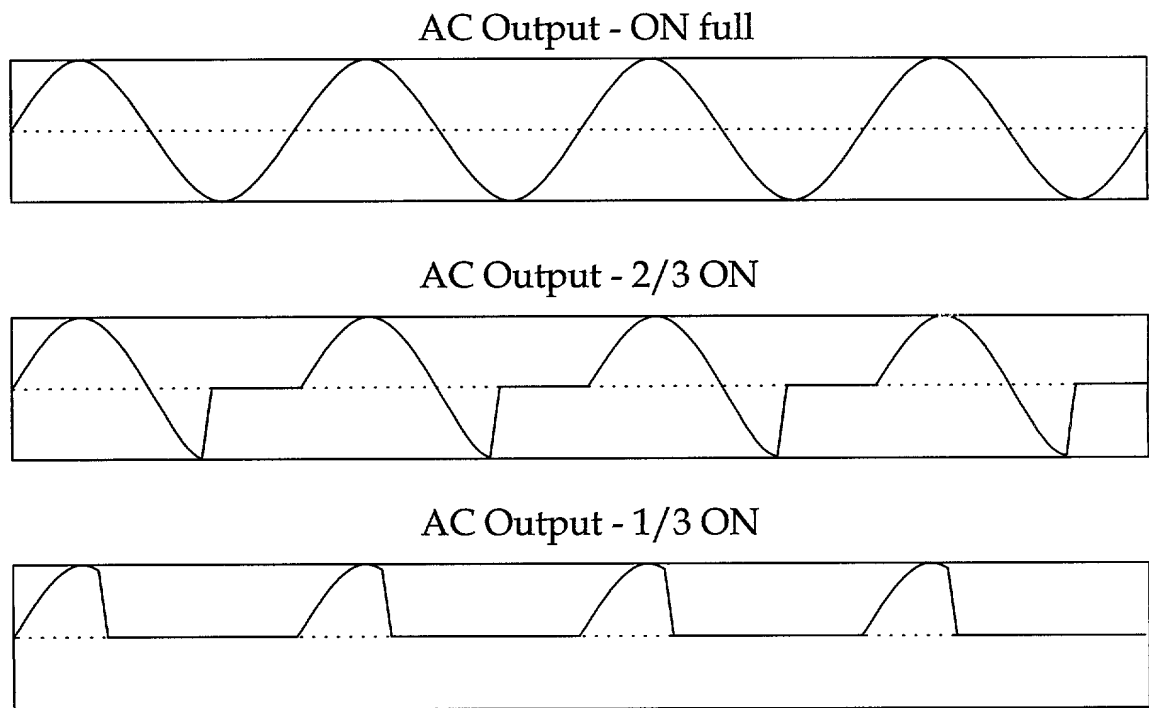


FIG. 2

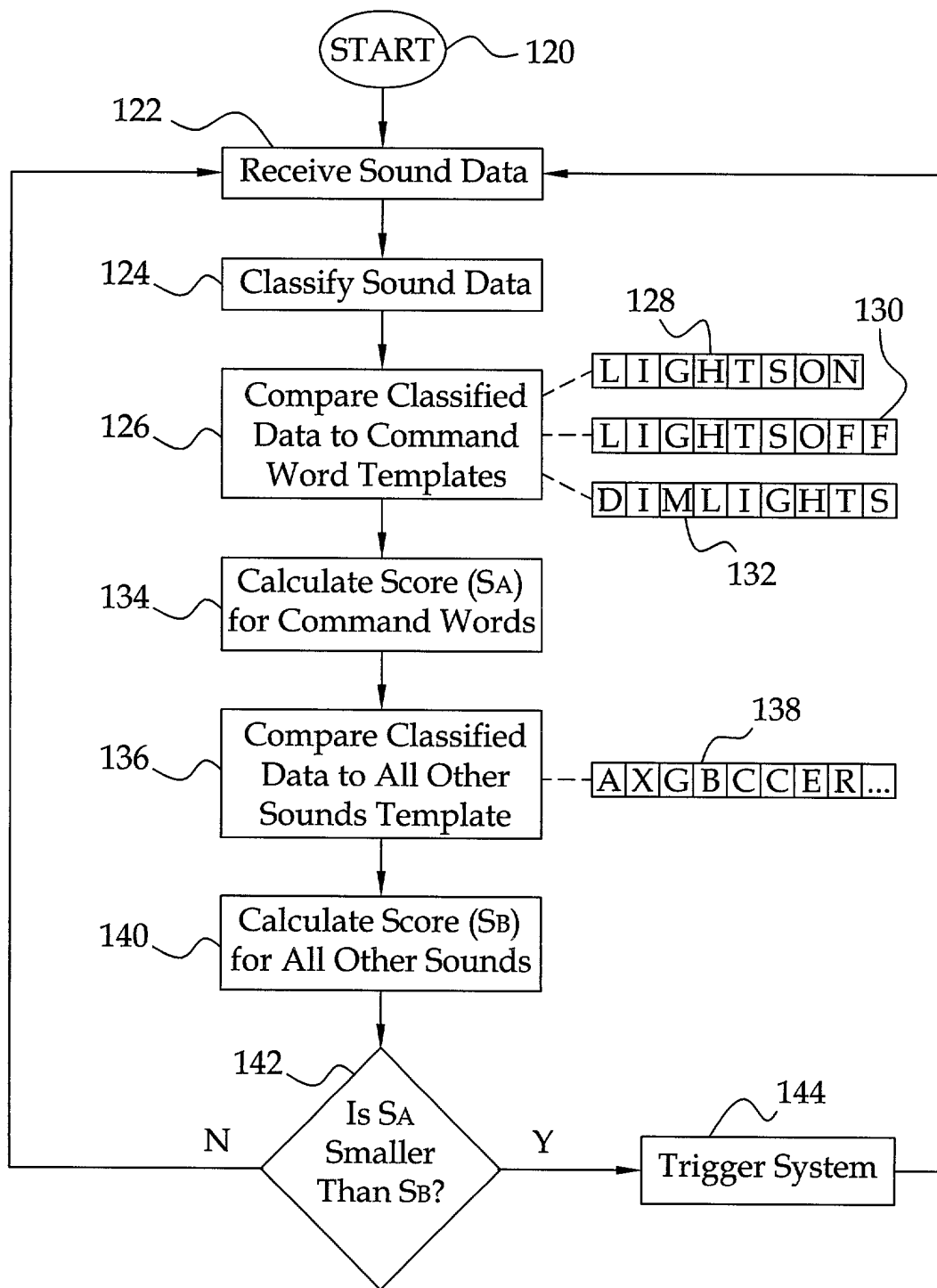


FIG. 3

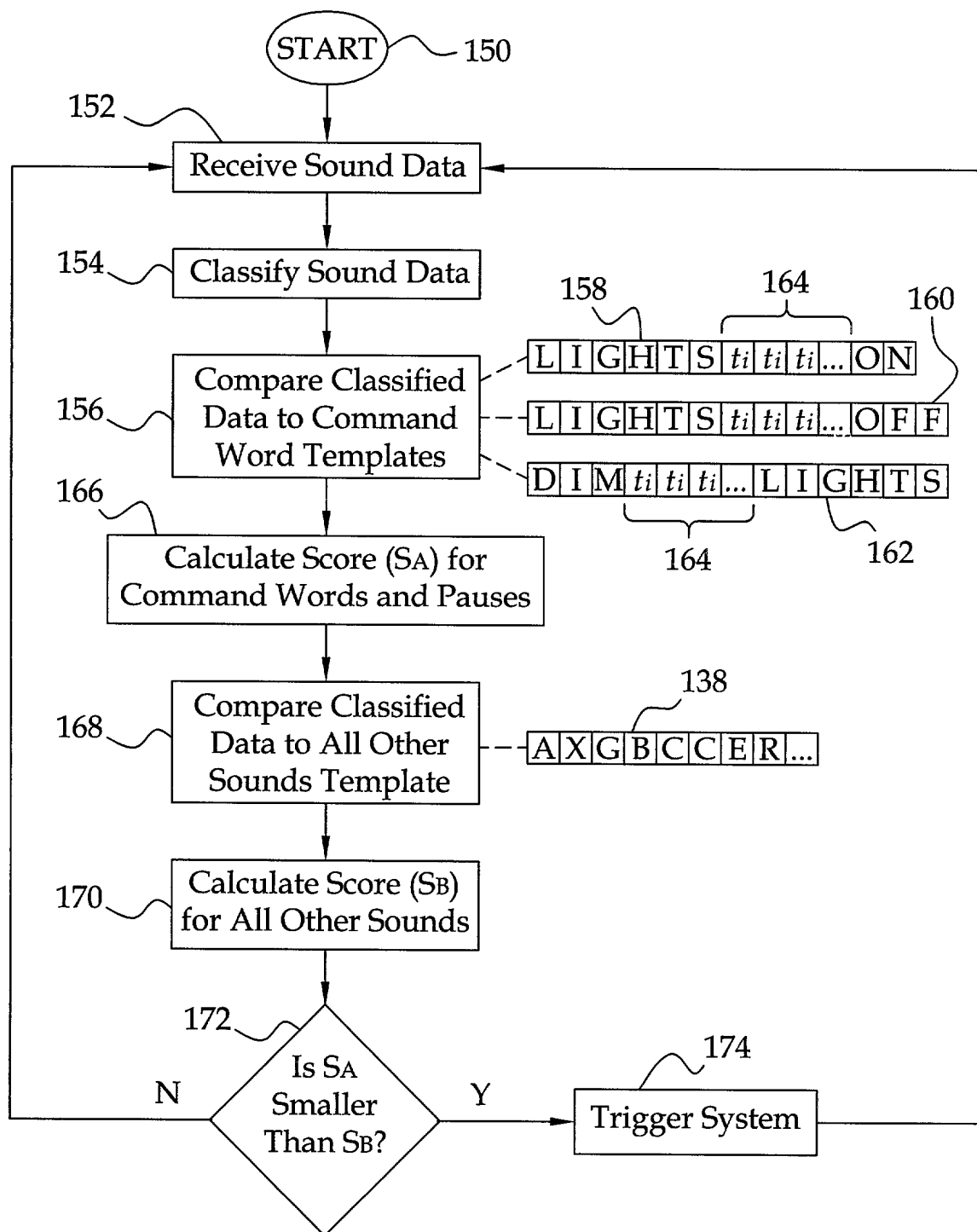


FIG. 4

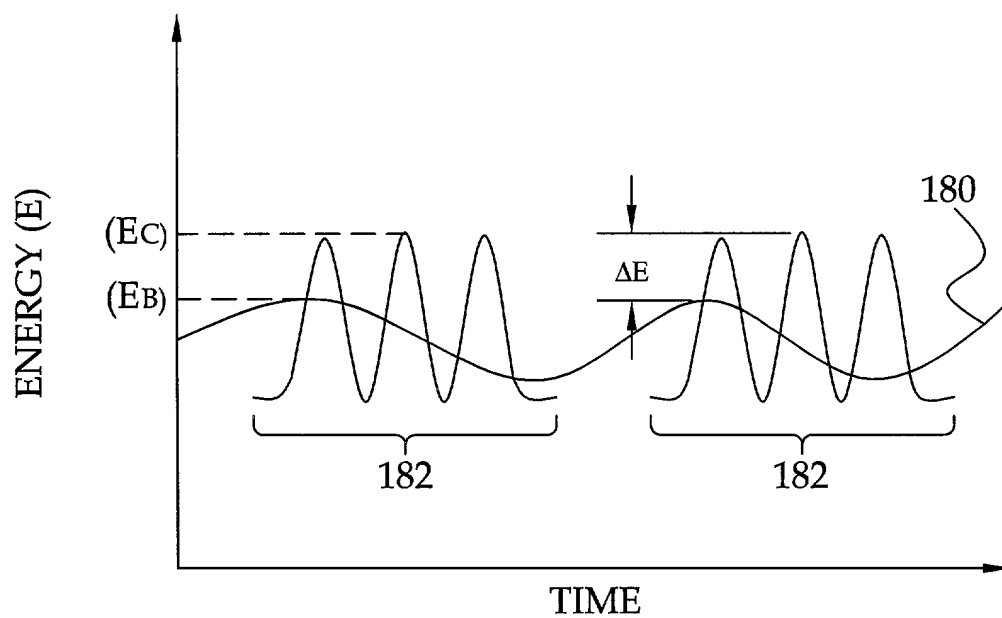


FIG. 5

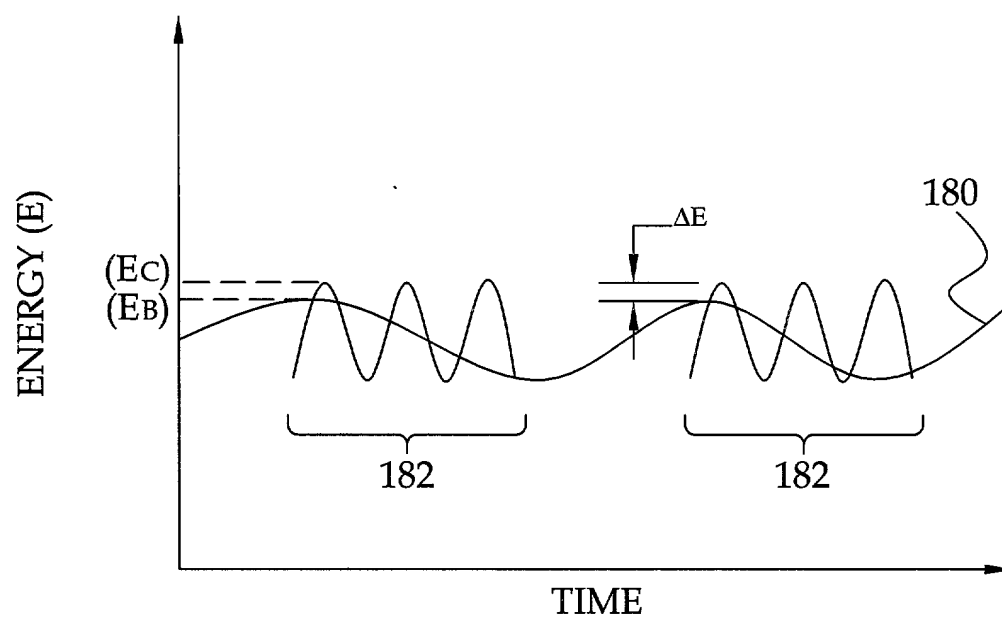


FIG. 6

# FIG. 7

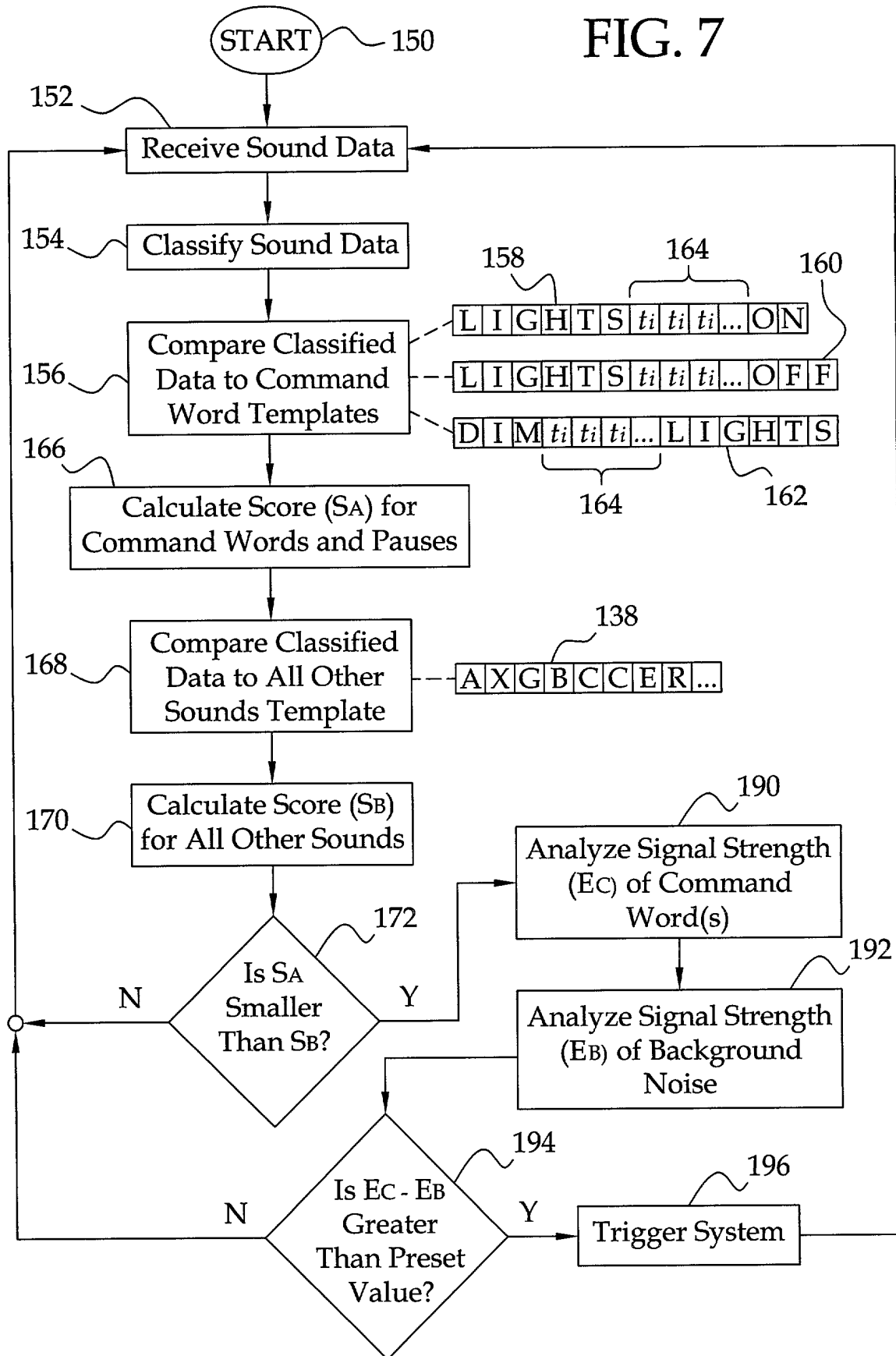
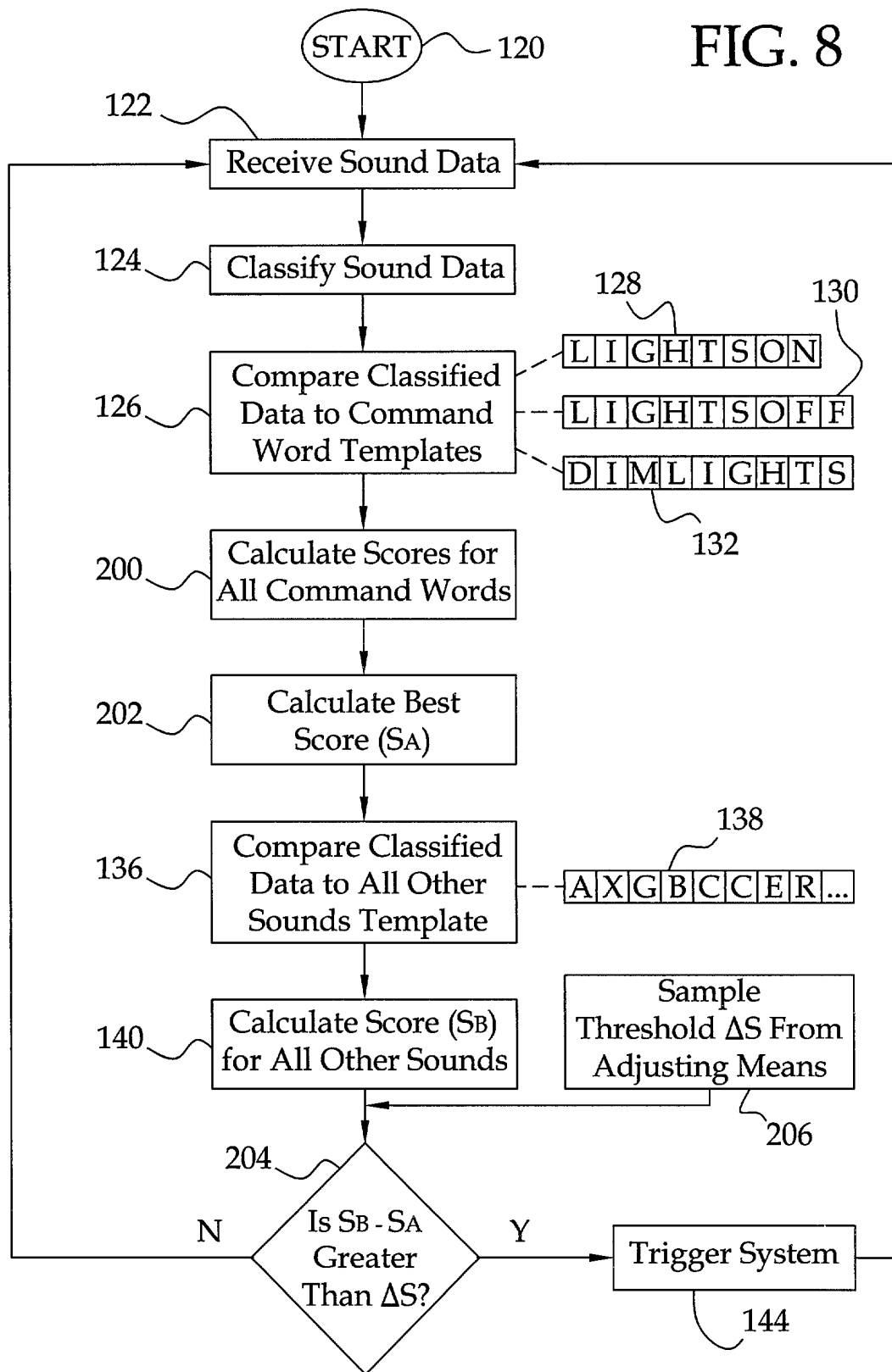


FIG. 8



---

**COMBINED DECLARATION AND POWER OF ATTORNEY**

(ORIGINAL, DESIGN, NATIONAL STAGE OF PCT, SUPPLEMENTAL, DIVISIONAL,  
CONTINUATION OR C-I-P)

---

As a below named inventor, I hereby declare that:

**TYPE OF DECLARATION**

This declaration is of the following type:

*(check one applicable item below)*

- ☒ original.
- ☐ design.
- ☐ supplemental.

NOTE: If the declaration is for an International Application being filed as a divisional, continuation or continuation-in-part application, do not check next item; check appropriate one of last three items.

- ☐ national stage of PCT.

NOTE: If one of the following 3 items apply, then complete and also attach ADDED PAGES FOR DIVISIONAL, CONTINUATION OR C-I-P.

- ☐ divisional.
- ☐ continuation.
- ☒ continuation-in-part (C-I-P).

**INVENTORSHIP IDENTIFICATION**

**WARNING:** If the inventors are each not the inventors of all the claims, an explanation of the facts, including the ownership of all the claims at the time the last claimed invention was made, should be submitted.

My residence, post office address and citizenship are as stated below, next to my name. I believe that I am the original, first and sole inventor *(if only one name is listed below)* or an original, first and joint inventor *(if plural names are listed below)* of the subject matter that is claimed, and for which a patent is sought on the invention entitled:

**TITLE OF INVENTION**

---

**Voice-Activated Control For Electrical Device**

---

## SPECIFICATION IDENTIFICATION

the specification of which:

(complete (a), (b) or (c))

(a) ☒ is attached hereto.

NOTE: "The following combinations of information supplied in an oath or declaration filed on the application filing date with a specification are acceptable as minimums for identifying a specification and compliance with any one of the items below will be accepted as complying with the identification requirement of 37 CFR 1.63:

"(1) name of inventor(s), and reference to an attached specification which is both attached to the oath or declaration at the time of execution and submitted with the oath or declaration on filing;

"(2) name of inventor(s), and attorney docket number which was on the specification as filed; or

"(3) name of inventor(s), and title which was on the specification as filed."

Notice of July 13, 1995 (1177 O.G. 60).

(b) ☐ was filed on \_\_\_\_\_, as ☐ Serial No. 0 / \_\_\_\_\_  
or ☐ \_\_\_\_\_  
and was amended on \_\_\_\_\_ (if applicable).

NOTE: Amendments filed after the original papers are deposited with the PTO that contain new matter are not accorded a filing date by being referred to in the declaration. Accordingly, the amendments involved are those filed with the application papers or, in the case of a supplemental declaration, are those amendments claiming matter not encompassed in the original statement of invention or claims. See 37 CFR 1.67.

NOTE: "The following combinations of information supplied in an oath or declaration filed after the filing date are acceptable as minimums for identifying a specification and compliance with any one of the items below will be accepted as complying with the identification requirement of 37 CFR 1.63:

"(1) name of inventor(s), and application number (consisting of the series code and the serial number; e.g., 08/123,456);

"(2) name of inventor(s), serial number and filing date;

"(3) name of inventor(s) and attorney docket number which was on the specification as filed;

"(4) name of inventor(s), title which was on the specification as filed and filing date;

"(5) name of inventor(s), title which was on the specification as filed and reference to an attached specification which is both attached to the oath or declaration at the time of execution and submitted with the oath or declaration; or

"(6) name of inventor(s), title which was on the specification as filed and accompanied by a cover letter accurately identifying the application for which it was intended by either the application number (consisting of the series code and the serial number; e.g., 08/123,456), or serial number and filing date. Absent any statement(s) to the contrary, it will be presumed that the application filed in the PTO is the application which the inventor(s) executed by signing the oath or declaration."

Notice of July 13, 1995 (1177 O.G. 60).

(c) ☐ was described and claimed in PCT International Application No. \_\_\_\_\_, filed on \_\_\_\_\_ and as amended under PCT Article 19 on \_\_\_\_\_ (if any).

## ACKNOWLEDGEMENT OF REVIEW OF PAPERS AND DUTY OF CANDOR

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information, which is material to patentability as defined in 37, Code of Federal Regulations, § 1.56,

*(also check the following items, if desired)*

- ☒ and which is material to the examination of this application, namely, information where there is a substantial likelihood that a reasonable Examiner would consider it important in deciding whether to allow the application to issue as a patent, and
- ☐ in compliance with this duty, there is attached an information disclosure statement, in accordance with 37 CFR 1.98.

### PRIORITY CLAIM (35 U.S.C. § 119(a)-(d))

I hereby claim foreign priority benefits under Title 35, United States Code, § 119(a)-(d) of any foreign application(s) for patent or inventor's certificate or of any PCT international application(s) designating at least one country other than the United States of America listed below and have also identified below any foreign application(s) for patent or inventor's certificate or any PCT international application(s) designating at least one country other than the United States of America filed by me on the same subject matter having a filing date before that of the application(s) of which priority is claimed.

*(complete (d) or (e))*

- (d) ☐ no such applications have been filed.
- (e) ☐ such applications have been filed as follows.

**NOTE:** Where item (c) is entered above and the International Application which designated the U.S. itself claimed priority check item (e), enter the details below and make the priority claim.

0044957-111000

**PRIOR FOREIGN/PCT APPLICATION(S) FILED WITHIN 12 MONTHS  
(6 MONTHS FOR DESIGN) PRIOR TO THIS APPLICATION  
AND ANY PRIORITY CLAIMS UNDER 35 U.S.C. § 119(a)-(d)**

COUNTRY (OR INDICATE IF PCT)	APPLICATION NUMBER	DATE OF FILING (day, month, year)	PRIORITY CLAIMED UNDER 37 USC 119
			<input type="checkbox"/> YES    NO <input type="checkbox"/>
			<input type="checkbox"/> YES    NO <input type="checkbox"/>
			<input type="checkbox"/> YES    NO <input type="checkbox"/>
			<input type="checkbox"/> YES    NO <input type="checkbox"/>
			<input type="checkbox"/> YES    NO <input type="checkbox"/>

**CLAIM FOR BENEFIT OF PRIOR U.S. PROVISIONAL APPLICATION(S)**  
(34 U.S.C. § 119(e))

I hereby claim the benefit under Title 35, United States Code, § 119(e) of any United States provisional application(s) listed below:

**PROVISIONAL APPLICATION NUMBER**

**FILING DATE**

\_\_\_\_ / \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_ / \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_ / \_\_\_\_\_

\_\_\_\_\_

**CLAIM FOR BENEFIT OF EARLIER US/PCT APPLICATION(S)**  
**UNDER 35 U.S.C. 120**

- ☒ The claim for the benefit of any such applications are set forth in the attached ADDED PAGES TO COMBINED DECLARATION AND POWER OF ATTORNEY FOR DIVISIONAL, CONTINUATION OR CONTINUATION-IN-PART (C-I-P) APPLICATION.

**NOTE:** If the application filed more than 12 months from the filing date of this application is a PCT filing forming the basis for this application entering the United States as (1) the national stage, or (2) a continuation, divisional, or continuation-in-part, then also complete ADDED PAGES TO COMBINED DECLARATION AND POWER OF ATTORNEY FOR DIVISIONAL, CONTINUATION OR C-I-P APPLICATION for benefit of the prior U.S. or PCT application(s) under 35 U.S.C. § 120.

*(list name and registration number)*

Lawrence G. Fridman	Siegmar Silber
Reg. No. 31,615	Reg. No. 26,233

☐ Attached, as part of this declaration and power of attorney, is the authorization of the above-named attorney(s) to accept and follow instructions from my representative(s).

Lawrence G. Fridman, Esq.  
Silber & Fridman  
66 Mount Prospect Ave.  
Clifton, N.J. 07013-1918

Lawrence G. Fridman, Esq.  
(973) 779-2580

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

(Declaration and Power of Attorney [1-1]—page 5 of 7)

### SIGNATURE(S)

NOTE: Carefully indicate the family (or last) name, as it should appear on the filing receipt and all other documents.

Full name of sole or first inventor

Igor (GIVEN NAME) NMN (MIDDLE INITIAL OR NAME) Zlozkarnik (FAMILY (OR LAST) NAME)  
Inventor's signature  
Date 11/15/99 Country of Citizenship Germany  
Residence 119 Lexington Street, Watertown, MA 02472  
Post Office Address Same as above

Full name of second joint inventor, if any

Daniel (GIVEN NAME) Lawrence (MIDDLE INITIAL OR NAME) Roth (FAMILY (OR LAST) NAME)  
Inventor's signature  
Date Nov 15 1999 Country of Citizenship USA  
Residence 10 Thacher Street, Boston, MA 02113  
Post Office Address Same as above

66611-1 25000000

(check proper box(es) for any of the following added page(s)  
that form a part of this declaration)

- ☐ **Signature** for fourth and subsequent joint inventors. *Number of pages added* \_\_\_\_\_

\* \* \*

- ☐ **Signature** by administrator(trix), executor(trix) or legal representative for deceased or incapacitated inventor. *Number of pages added* \_\_\_\_\_

\* \* \*

- ☐ **Signature** for inventor who refuses to sign or cannot be reached by person authorized under 37 CFR 1.47. *Number of pages added* \_\_\_\_\_

\* \* \*

- ☐ Added page for **signature** by one joint inventor on behalf of deceased inventor(s) where legal representative cannot be appointed in time. (37 CFR 1.47)

\* \* \*

- ☐ Added pages to combined declaration and power of attorney for divisional, continuation, or continuation-in-part (C-I-P) application.

☐ Number of pages added \_\_\_\_\_

\* \* \*

- ☐ Authorization of attorney(s) to accept and follow instructions from representative.

\* \* \*

(If no further pages form a part of this Declaration,  
then end this Declaration with this page and check the following item)

- ☒ This declaration ends with this page.

09413957-11099  
000011/20061100